

MATE-T580 Practical Data Science using R

Assignment 3

Visualizing World Healthcare Data

In assignment 1, we compared healthcare outcomes of developed countries based on OECD data. In this assignment, we come back to the same datasets. You are going to assume the role of a **data journalist**. Data journalists tell a story by producing a high quality visual from data. Thus, your task is to produce one revealing visualization from the healthcare data using the ggplot2 library. This is an open-ended task. You are not to limit yourself to the same analysis done in assignment 1. Example, you may consider the time dimension, other healthcare expenditure metrics, other healthcare outcome metrics, comparison between males and females, etc. Be creative! Given the richness of the datasets, I expect that no two students shall consider the same data dimensions in their plots. Here are some guiding principles that should help you:

- The plot should not be too simple (dimension-wise) or boring. Example, a bar plot ranking countries according to some metric is not considered acceptable.
- You will have to do some data manipulations using the dplyr/tidyr libraries to prepare a dataframe suitable to pass to ggplot.
- The plot should be clear and should tell a story on its own (pay attention to units, title, caption, axes labels, axes scales, legends, etc.).
- You may (but don't have to) use both datasets (expenditures and outcomes). Each dataset is rich enough to pull interesting information from.
- You may (but don't have to) use geometries we haven't discussed in class (e.g. geom_path, geom_line, geom_map, etc.). ggplot has many other geoms.
- The only line of code I need to change to run your script should be the setwd function that points to the working folder where the csv files are found.

Appendix

The csv files can be downloaded from:

<https://drive.google.com/uc?export=download&id=1yKjNfEtjmNYJBXLUFcJtzkaIrPElzZT7>

<https://drive.google.com/uc?export=download&id=1VSoN8CVRsQKluNAS8lNYPxJAeeEo7bV->